Comparing physical activity estimates in children from hip-worn Actigraph GT3X+ accelerometers using raw and counts based processing methods

Buchan, Duncan S.; McLellan, Gillian

Published in:
Journal of Sports Sciences

DOI:
10.1080/02640414.2018.1527198

E-pub ahead of print: 12/10/2018

Document Version
Peer reviewed version

Citation for published version (APA):
Buchan, D. S., & McLellan, G. (2018). Comparing physical activity estimates in children from hip-worn Actigraph GT3X+ accelerometers using raw and counts based processing methods. *Journal of Sports Sciences*, *37*(7), 779-787. https://doi.org/10.1080/02640414.2018.1527198

Comparing Physical Activity estimates in children from hip-worn Actigraph GT3X+ accelerometers using raw and counts based processing methods

Duncan S. Buchan* and Gillian McLellan

Department of Sport and Exercise, The University of the West of Scotland, Hamilton, United Kingdom.

* Corresponding author

Email: duncan.buchan@uws.ac.uk

Dr Duncan Buchan
Senior Lecturer in Exercise Science
Institute of Clinical Exercise and Health Science
School of Science and Sport
University of the West of Scotland
Almada Street
Hamilton (Room A430)
South Lanarkshire
United Kingdom
ML3 0JB

Abstract:

This study examined differences in physical activity (PA) estimates provided from raw and counts processing methods. One hundred and sixty-five children (87 girls) wore a hip-mounted ActiGraph GT3X+ accelerometer for 7 days. Data were available for 129 participants. Time in moderate PA (MPA), vigorous PA (VPA) and moderate-vigorous PA (MVPA) were calculated using R-package GGIR and ActiLife. Participants meeting the wear time criteria for both processing methods were included in the analysis. Time spent in MPA (-21.4 $min.d^{-1}$, 95%CI -21 to -20) and VPA (-36 $min.d^{-1}$, 95%CI -40 to -33) from count data were higher (*P*<0.001) than raw data. Time spent in MVPA between the two processing methods revealed significant differences (All *P*<0.001). Bland-Altman plots suggest that the mean bias for time spent in MPA, VPA and MVPA were large when comparing raw and count methods. Equivalence tests showed that estimates from raw and count processing methods across all activity intensities lacked equivalence. Lack of equivalence and poor agreement between raw and count processing methods suggest the two approaches to estimate PA are not comparable. Further work to facilitate the comparison of findings between studies that process and report raw and count physical activity data may be necessary.

Keywords: ActiGraph, GGIR, GT3X+, Physical Activity, Youth, Accelerometers.

Acknowledgements

The authors would like to thank the children, schools, teachers and parents who agreed to participate in this study. This study was funded by The University of the West of Scotland, VP Research Fund. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. We would like to thank Dr Lynne Boddy for her assistance with the Equivalence analysis figures.

Authors' Contributions

GMcL coordinated the study, collected all the data and provided comments on the draft manuscript. DSB conceived and designed the study, performed the statistical analysis and wrote the manuscript. Both authors have read and approved the final version of the manuscript and agree with the order of presentation of authors.

None of the authors declare competing financial interests.

Conflicts of interest: none.

Introduction

The most widely used objective measure of youth physical activity (PA) patterns is through the use of accelerometers (Cain, Sallis, Conway, Van Dyck, & Calhoon, 2013). Historically, accelerometers have been worn at the hip to capture accelerations of whole body movements. This raw acceleration data is subsequently converted into proprietary counts which are then used to estimate time spent in activity intensities based on published thresholds or cut-points (Cain et al., 2013). An important technological advancement in accelerometery has been the ability to access the triaxial raw acceleration data prior to being processed, filtered and scaled from devices such as the ActiGraph GT3X+ (Pensacola, FL, USA), GENEActiv (ActivInsights Ltd., Cambridge, UK) and Axivity (Axivity Ltd, Newcastle, UK). This move towards raw data processing affords greater transparency and consistency to post-data processing methodologies and greater measurement opportunities but may come at a cost (Rowlands et al., 2016).

There is a wealth of accelerometer data that has been collected and analysed over the years using the counts based approach (Cooper et al., 2015; Katzmarzyk et al., 2015). As of 2018, more than 37,000 young people aged 3 to 18 years across studies from Europe, the US, Brazil and Australia have provided PA data from hip-worn ActiGraphs in the International Children's Accelerometry Database (ICAD) (Cooper et al., 2015). More recently the International Study of Childhood Obesity, Lifestyle and the Environment (ISCOLE) collected PA data on 6000 children aged 9-11 years across five diverse geographic regions of the globe using hip-worn ActiGraphs (Katzmarzyk et al., 2015). As the field of accelerometery moves towards the use of raw data processing techniques, studies that examine the comparability of PA outcomes derived from raw and count processing methods are needed.

The ActiGraph accelerometers are commonly used by researchers owing to the large body of evidence supporting its use (Cain et al., 2013). To the best of our knowledge however, only one study has examined the comparability of activity outputs from hip-mounted ActiGraph GT3X+ raw and count processing methods in children (Fairclough et al., 2016). Here the authors examined differences in PA estimates derived from Euclidean Norm Minus One (ENMO) and count data using the cut-points proposed by Evenson et al. (Evenson, Catellier, Gill, Ondrak, & McMurray, 2008). Findings revealed that estimates of time spent in moderate PA (MPA) from ENMO was greater than when processing the count data. Conversely, time spent in vigorous PA (VPA) was seen to be lower when using ENMO compared to count data. There has been some consensus in recent years for the use of the Evenson et al. cut-points based on the convincing evidence of their validity in children (Trost, Loprinzi, Moore, & Pfeiffer, 2011). Yet, different generations of ActiGraph devices are known to provide contrasting estimates of PA in a free-living environment (Grydeland, Hansen, Ried-Larsen, Kolle, & Anderssen, 2014; Ried-Larsen et al., 2012). This suggests that cut-points developed using an older generation of the ActiGraph accelerometer may not be used for data collected with more recent devices such as the GT3X+. Since the Evenson cut-points were developed using the ActiGraph 7164 uniaxial accelerometer and the Trost validation study was undertaken using the ActiGraph GM1 accelerometer (Evenson et al., 2008; Trost et al., 2011), it is unclear whether the magnitude of differences reported by Fairclough et al. would differ if cut-points developed for use with the GT3X+ accelerometer were used.

Advances in accelerometer technology saw the release of the GT3X+ model in 2010 which has an increased capacity for data storage and the ability to measure accelerations across three axes, unlike the previous GM1 device. Since the GT3X+ can measure accelerations in three individual plains of motion (i.e. vertical, antero-posterior and medio-lateral), which can be examined either individually or together (vector magnitude (VM)), in theory, they should provide a more accurate assessment of PA when compared to a single axis that is capable of measuring acceleration only in the vertical direction. Indeed, some studies have reported improved energy expenditure estimates based on VM counts when compared to vertical counts alone (Eston, Rowlands, & Ingledew, 1998; Rowlands, Thomas, Eston, & Topping, 2004), although it should be acknowledged there is evidence which refutes these findings (Kavouras, Sarras, Tsekouras, & Sidossis, 2008).

In order to capitalise on the these new features within triaxial accelerometers, researchers have developed VM cut-points for use with the GT3X+ accelerometer (Hänggi, Phillips, & Rowlands, 2013; Romanzini, Petroski, Ohara, Dourado, &

Reichert, 2014). In these laboratory studies, oxygen uptake on a breath-by-breath basis is measured using a portable metabolic system as participants perform a series of predefined activities that provide a range of energy expenditure responses (Trost et al., 2011). This data is then transformed into age-specific metabolic equivalents which is matched with the acceleration data to define an acceleration count range which constitutes sedentary, light, moderate and vigorous intensity activities.

With the availability of several cut-points to process accelerometer counts data, the choice of which cut-points to use can have a significant impact upon activity estimates and study comparability (Logan, Duncan, Harris, Hinckson, & Schofield, 2016; Pate et al., 2006). Yet, no study has assessed the comparability of outputs from raw and counts based processing methods using cut-points specifically developed for the triaxial ActiGraph GT3X+ (Hänggi et al., 2013; Romanzini et al., 2014; Santos-Lozano et al., 2013). Given the increased availability of triaxial accelerometers and available VM cut-points in children and adolescents, it is important to establish the comparability of VM counts output to that of other processing methods. Thereafter, findings could be used to determine whether it is necessary to develop a data reduction procedure, as proposed to harmonize estimates of MVPA derived from wrist- and hip-worn accelerometers (Brazendale et al., 2018), to enhance the comparability of estimates between different processing methods. Therefore, the purpose of this study is to compare activity estimates derived from raw and counts based processing methods from a hip-mounted GT3X+ accelerometer.

Methods

Participants included 165 children (87 girls) aged 10.4 ± 0.9 years from three schools in South Lanarkshire, Scotland. Three schools were provided with 70 information and consent forms (n=210) to be distributed to children from years 5-7. Written consent and assent was provided before children participated in the study. Ethical approval was provided through the University of the West of Scotland's ethical committee with data collection taking place between September and December 2016.

Anthropometrics

Stature was measured barefoot to the nearest 0.1cm using a portable stadiometer (Seca Stadiometer, Seca Ltd, Birmingham, UK). Mass was measured barefoot with light clothing to the nearest 0.1kg on electronic scales (Seca Digital Scales, Seca Ltd, Birmingham, UK). From measured stature and mass, BMI-z scores were calculated for each participant relative to the UK 1990 BMI population reference data using software provided by the Child Growth Foundation (Cole, Freeman, & Preece, 1995; Pan & Cole, 2010).

Assessment of activity

Free-living activity was captured using the ActiGraph GT3X+ positioned above the right hip on a belt worn around the waist. Prior to distribution, the accelerometer was synchronised with Greenwich Mean time, initialized to capture data at 80Hz and programmed to commence data collection at 6:00am on the day following participants receiving the devices. As the aim of this study was not to examine light intensity physical activity or sedentary behaviour, the low frequency extension was not enabled. Participants were instructed to wear the accelerometer at all times (i.e. 24 hours) for seven days, apart from water-based activities such as swimming or bathing. The requirement to wear the hip-worn accelerometer for 24 hours is in line with previous studies who reported an average wear time of 22.6 $h.d^{-1}$ from hip-worn ActiGraph GT3X+ accelerometers and a waking wear time of 14.7 $h.d^{-1}$ (Tudor-Locke et al., 2015). Since poor compliance and subsequent selection bias and misclassification is often cited as a limitation of hip-worn accelerometer studies (Troiano, McClain, Brychta, & Chen, 2014), we used the 24 hour wear time protocol to encourage compliance. To ensure the correct placement of the device, participants were fitted with the accelerometer prior to leaving the testing session.

Data management

Upon the return of the device, data were uploaded using ActiLife v6.13.3 (Actigraph, Pensacola, FL, USA) and saved in raw format as GT3X+ files. The GT3X+ files were used to generate 1 s epoch csv files containing x, y and z vectors to facilitate raw data processing and to AGD format to facilitate the analysis of count data. The 1 s epoch csv files were then processed in R (http://cran.r-project.org) using the GGIR package (version 1.5-10) which allows raw accelerations (gravitational acceleration) to be processed and analysed (Van Hees et al., 2014). The GGIR package autocalibrates the raw triaxial accelerometer signals which subsequently converts the raw triaxial accelerometer data into one omnidirectional measure of acceleration, termed the signal vector magnitude (SVM). Further details of the autocalibration method has been described elsewhere (Van Hees et al., 2014). SVM was calculated from raw accelerations from the three axes minus 1 g, which represents the value of gravity (i.e., SVM = $\sqrt{(x^2 + y^2 + z^2)}$-1), with negative values rounded to zero. This metric has been referred to as the ENMO previously (Fairclough et al., 2016; Van Hees et al., 2014). Raw data were further reduced by calculating the average SVM values per 1 s epoch expressed in mg over each of the monitored days. Thereafter, raw data wear times were estimated on the basis of the standard deviation and value range of each axis, calculated for 60 min windows with 15-min moving increments as described in detail elsewhere (van Hees et al., 2013). The default setting for nonwear was used whereby invalid data were imputed by the average at similar time points on different days of the week. Raw files were removed from the analysis if post calibration error (deviation from 1 g during no movement) was greater than 0.02 g, as previously reported (Rowlands, Cliff, et al., 2016; Rowlands, Yates, Davies, Khunti, & Edwardson, 2016). For the GT3X+ count data, nonwear time periods were captured using the algorithms of Choi et al. and removed from the analysis to remain consistent with previous studies (Choi, Liu, Matthews, & Buchowski, 2011; Fairclough et al., 2016). Finally, sleep time (11:00pm to 6:00am) was removed to facilitate comparisons between the raw and count processing methods.

Data processing

To the best of our knowledge only one study has provided cut-points for ENMO to classify MVPA from hip-worn accelerometers (Hildebrand, Van Hees, Hansen, & Ekelund, 2014). Therefore, raw acceleration data were analysed using the device specific prediction equations provided by Hildebrand et al. to generate intensity specific milli-g cut-points based on the ENMO metric to classify minutes of MPA, VPA and MVPA (Hildebrand et al., 2014). These cut-points were: 142.6 - 464.5 mg

(MPA) and ≥464.6 mg (VPA). For the analysis of VM count based activity, we are aware of only three studies which have provided cut-points to classify MVPA from hip-worn accelerometers in youths (Hänggi et al., 2013; Romanzini et al., 2014; Santos-Lozano et al., 2013). Therefore, minutes of MPA, VPA and MVPA were calculated according to the cut-points provided by Romanzini et al. (Romanzini et al., 2014). These cut-points were: 757 - 1111 counts per 15 s (MPA) and ≥1112 counts per 15 s (VPA). Only minutes of MVPA were calculated according to the cut-points (≥56 counts per s) provided by Hänggi et al. since the authors did not provide VPA cut-points (Hänggi et al., 2013). Since the use of 60-second epochs may obscure short bursts of VPA and underestimate the activity profiles of children (Sanders, Cliff, & Lonsdale, 2014), we decided not to include the cut-points proposed by Santos-Lozano et al. (Santos-Lozano et al., 2013) within this study. Finally, as the cut-points provided by Hänggi et al. and Romanzini et al were developed using short epochs (1s and 15-secs), the GT3X+ files were downloaded as 1s and 15s epoch AGD files for the analysis of count data which was cleaned and scored using ActiLife v6.13.3.

To ensure as fair a comparison as possible between outputs, only participants that met the wear-time criteria for each method were included in the subsequent analysis. The wear-time criteria for participants was ≥ 10hours/day on at least one or more days as this is more likely to reflect the true PA and sedentary behaviour patterns of participants over the course of a day (Migueles et al., 2017). Any days with fewer than 10hours/day of wear time were removed from the analysis. In keeping with previous studies, participants with at least one day or more wear time were included as the purpose of this study was not to establish estimates of habitual PA, but rather compare outputs between raw and count based approaches (Rowlands et al., 2014). Nonetheless, we have reported the number of participants who met the minimal requirements of 60 minutes of MVPA per day as a guide and to assist with study comparisons.

Analysis

Descriptive characteristics were calculated for all output variables. To assess the comparability between the minutes accumulated at each intensity, several correlations and repeated measures ANOVA’s were undertaken. Correlations were undertaken for mean minutes accumulated at each intensity whereas ANOVA's were undertaken for daily minutes accumulated at each activity intensity. Initial analysis suggested that sex and age had no effect upon the agreement between minutes accumulated at each intensity, so both variables were removed from the analysis. Post-hoc analyses were undertaken using pairwise comparisons. Bland-Altman procedures (Altman & Bland, 1983) were used to assess agreement and systematic bias at the individual level between time estimates in each activity derived from raw and counts based approaches. Finally, the equivalence of time estimates between devices for time spent in each activity were examined at the group level using the 95% paired equivalence test. To reject the null-hypothesis, the 90% confidence interval (CI) for the mean derived from either raw or counts based approaches had to fall within an equivalence region defined as ±10% of the chosen reference method. In the absence of an empirically derived range for the equivalence zone, our decision to select ±10% of the mean as the proposed equivalence zone follows that of others (Boddy et al., 2018; Rowlands et al., 2017). The use of equivalence testing is increasing within PA related research when researchers are interested in knowing whether outcomes are statistically equivalent at the group level, rather than whether they are statistically different (Boddy et al., 2018; Dixon et al., 2018; Rowlands et al., 2017). To address the aims of this study, it was useful to know whether outcomes were statistically equivalent and processing methods provided similar estimates. Statistical significance was set at $P < 0.05$. Data are presented as mean ± SD unless otherwise stated.

Results

Although 165 children agreed to participate, there were 4 device malfunctions which resulted in no data being captured for these children. Thirty-five children (15 girls), who were significantly older ($P<0.05$) than included participants, failed to meet the wear time criteria in both ActiLife v6.13.3 and in GGIR and were subsequently removed from the analysis. No participant met the wear time criteria of one method and not the other. No data files were excluded based on calibration error when analysing the raw data, which resulted in a final sample of 129 children (71 girls; age 9.1 ± 0.9 years; stature 145.1 ± 8.6 cm; mass 40.1 ± 11.3 kg; BMI-z score 1.1 ± 1.0). Time spent in each activity intensity across processing methods are displayed in

Table 1. Analysis of time spent in each activity from raw and count data using the Romanzini et al. and Hildebrand et al. cut-points revealed that MPA (r=0.63), VPA (r=0.66) and MVPA (r=0.75) were moderately to strongly correlated (*P*<0.001). Analysis of time spent in MVPA between the raw and count data using the Hänggi et al. cut-points were strongly correlated (r=0.77, *P*<0.001). Finally, comparisons between the count data revealed that time spent in MVPA were strongly correlated (r=0.91, *P*<0.001).

Figure 1 displays the comparisons between time spent in activity intensities from the different processing methods. Time spent in MPA (-21.4 min.d^{-1}, 95%CI -21 to -20 min.d^{-1}) and VPA (-36 min.d^{-1}, 95%CI -40 to -33 min.d^{-1}) was significantly less (both *P*<0.001) using the Hildebrand et al. cut-points compared to the Romanzini et al. cut-points. Comparisons of time spent in MVPA between the raw and count processing methods revealed significant differences (All *P*<0.001). Less time in MVPA (-57.3 min.d^{-1}, 95%CI -62 to -53 min.d^{-1}) was observed using the Hildebrand et al. cut-points compared to the Romanzini et al. cut-points. Similarly, less time in MVPA (-65.6 min.d^{-1}, 95%CI -70 to -62 min.d^{-1}) was observed using the Hildebrand et al. cut-points compared to the Hänggi et al. cut-points. Less time in MVPA was also observed when using the Romanzini et al. cut-points compared to the Hänggi et al. cut-points (-8.3 min.d^{-1}, 95%CI -12 to -5 min.d^{-1}). Finally, the recommended 60 min.d^{-1} of MVPA was achieved by 86%, 95% and 8% of participants using the Romanzini et al., Hänggi et al. and Hildebrand et al. the cut-points.

Bland-Altman plots displayed in Figure 2A-E illustrate the degree of difference in time spent in MPA, VPA and MVPA between the processing methods. The mean bias for time spent in MPA and VPA were large as was the time spent in MVPA when comparing the raw and count processing methods (Figure 2 C-D). At the individual level, limits of agreement (LoA) were wide for time spent in MPA and VPA (Figure 2 A-B). The mean bias for time spent in MVPA when compared using count-based approaches was small (Figure 2E). At the individual level, the LoA were large which appeared to decrease with participants increasing level of PA engagement. Results of the equivalence testing can be found in Figure 3. Time spent in each activity intensity from the processing methods with 90% CIs can be found in Table 1. None of the 90% CIs for the Hildebrand et al. estimates of time spent in

MPA, VPA or MVPA fell within the zone of equivalence for the Romanzini et al. estimates. Nor did the 90%CIs of the Hänggi et al. estimates for time spent in MVPA when compared to the zone of equivalence for the Hildebrand et al. estimates. These finding suggest there were no statistically significant equivalence between the cut-points compared. Re-running the equivalence analyses with the alternate cut-points selected as the reference method did not change these interpretations. As the 90%CIs for the Hänggi et al. estimates for time spent in MVPA extend beyond the equivalence zone for the Romanzini et al. estimates, the different cut-points don't provide equivalent estimates on average.

Discussion

This study assessed the comparability of activity estimates from hip-mounted ActiGraph GT3X+ accelerometers using raw and count processing methods. Our findings suggest there are large differences in time spent in MPA, VPA and MVPA when comparing raw and count processing methods. Furthermore, our findings show that activity estimates lacked equivalence and demonstrated poor agreement when compared across processing methods. The lack of agreement in these methods for estimating PA in youth suggest that conclusions concerning youth activity patterns are influenced by the type of data being processed.

With the recent advancements made in accelerometer technology and the emergence of triaxial accelerometers, there have been calls for the use of triaxial cut-points to be used in future studies when providing estimates of youth activity (Kim et al., 2017; Logan et al., 2016). Research has shown that measuring accelerations over three axes, rather than one, can be more sensitive to movements and lead to higher estimates of activity (Logan et al., 2016; Santos-Lozano et al., 2013). For instance, Logan et al., reported 51.5%, 49.4%, 48.3%, 47.7% and 47.2% differences for total mean counts per day between VM and the vertical axis output for 1, 5, 15, 30 and 60 s epoch lengths, respectively (Logan et al., 2016). The authors also noted significant differences in time spent in MVPA between vertical axis and VM (Hänggi et al., 2013; Romanzini et al., 2014) cut-points which suggests that certain movements and accelerations could be missed by one axis but captured by the VM.

Moreover, previous studies comparing hip-worn GT3X VM estimates with that of criterion measures suggest they perform better than methods which rely upon the vertical axis (Crouter, Horton, & Bassett, 2013; Kim et al., 2016). Recent findings also suggest that associations between MVPA, estimated from ENMO, successful ageing (Menai et al., 2017) and adiposity markers (Sabia et al., 2015) is more pronounced when compared to questionnaires. Although these studies employed a wrist-worn accelerometer in their design, their findings do provide confidence to researchers of the performance of ENMO to identify well-established relationships. It is disappointing therefore to find such differences in activity estimates between the two VM methods (counts approach) and that of the GGIR method (raw approach) which limit the pooling of data from studies using these different accelerometer processing methods.

To the best of our knowledge, only one study has compared activity estimates from hip-mounted ActiGraph GT3X+ raw and counts data (Fairclough et al., 2016). Here the authors noted that time spent in MPA calculated from ENMO was 42.0 ± 1.6 min.d^{-1} compared to 35.1 ± 1.0 min.d^{-1} from counts data, a significant difference of 16.5%. The authors also reported that time spent in VPA significantly differed by 79.5% between count (37.1 ± 1.9 min.d^{-1}) and raw (7.6 ± 0.5 min.d^{-1}) processing methods. In a similar study by Rowlands and colleagues (Rowlands et al., 2014), the authors compared ActiGraph GT3X+ count data with that of the GENEActiv raw data when both devices were worn at the hip. Findings were largely similar to that of Fairclough et al. where time spent in MPA from raw data was greater than count data (56.7 ± 18.0 vs. 32.3 ± 9.2 min.d^{-1}), but lower for VPA (11.1 ± 7.0 vs. 30.0 ± 12.2 min.d^{-1}). Contrary to these findings, we found that time spent in MPA from raw data were lower than count data in this study (27.8 ± 11.1 vs. 49.2 ± 12.8 min.d^{-1}), as was time spent in VPA (6.6 ± 4.0 vs. 42.8 ± 21.8 min.d^{-1}). These combined differences reflect the overall MVPA differences when using the Romanzini et al. cut-points, with raw MVPA estimates being lower than when using the count data (35.0 ± 14.4 vs. 91.8 ± 31.2 min.d^{-1}). Similar differences were also evident when comparing the Hildebrand et al. cut-points to the Hänggi et al. cut-points to estimate time in MVPA (35.0 ± 14.4 vs. 100.1 ± 27.5 min.d^{-1}).

We observed significant differences in time spent in MVPA between the Hildebrand et al. cut-points and that of the counts cut-points (Figure 1) unlike Fairclough et al. Using the counts processing methods, we found that time spent in MVPA was 91.8 ± 31.2 min.d-1 using the Romanzini et al. cut-points and 100.1 ± 27.5 min.d-1 using the Hänggi et al. cut-points. This is in contrast to the time spent in MVPA (72.1 ± 2.6 min.d-1) reported by Fairclough et al. when using the vertical axis cut-points proposed by Evenson et al. We also found that time spent in MVPA from the raw data was lower than those reported by Fairclough et al. (34.5 ± 14.4 vs. 49.6 ± 2.0 min.d-1). Given these differences it is unsurprising we observed significant differences in time spent in MVPA between the raw and counts processing methods. In addition to the processing methods, the different populations and protocols used to generate cut-points is also a likely explanation for the differences in outcomes. For instance, Evenson et al. used children aged 5 to 8 years whereas Hänggi et al. and Romanzini et al. used children aged 10 to 15 years. Moreover, the type and number of physical activities undertaken in these studies were not consistent and were performed in durations of 7 mins (Evenson et al., 2008), 3 mins (Hänggi et al., 2013) and 5 mins (Romanzini et al., 2014).

As time spent in MVPA is often used to quantify the number of youth meeting current PA recommendations (Department of Health, 2011; US Department of Health and Human Services, 2008; World Health Organization, 2010), it is important that those in public health positions are aware that different processing approaches to accelerometer data may yield contrasting conclusions when commenting upon youth behaviour. To highlight this, in a recent study examining youth PA levels from wrist accelerometery, Kim et al. (Kim et al., 2017) found in a sample of 408 youths that 0% of participants achieved ≥ 60 $min.d^{-1}$ of MVPA when using the GGIR method to process raw acceleration data. When they examined activity levels using counts based approaches, the proportion of participants achieving ≥ 60 $min.d^{-1}$ of MVPA were substantially higher when using the Crouter (range, 43.5 – 69.9%) (Crouter et al., 2013) and Chandler (range, 6.2 – 23.2%) (Chandler, Brazendale, Beets, & Mealing, 2016) cut-points. The authors also noted that the proportion of youth achieving ≥ 60 $min.d^{-1}$ of MVPA were substantially higher when using VM rather than vertical axis cut-points. Although the accelerometers were placed on the wrist (Kim et al., 2017), the discrepancies in outcomes when processing accelerometer data using raw and counts based approaches are in line with the findings reported here. For instance, 86%, 95% and 8% of participants achieved the recommended 60 min.d^{-1} of MVPA when using the cut-points provided by Romanzini et al, Hänggi et al. and Hildebrand et al. whereas Fairclough et al. found that 20 (20%) and 56 (68%) participants with valid raw and counts data achieved the recommended 60 min.d^{-1} of MVPA.

The lack of equivalence and poor agreement between activity estimates when comparing raw and counts based approaches suggest that outcomes derived from such methods are not comparable. A recent study highlighted the poor classification performance of the Hildebrand cut-points for correctly classifying MVPA, primarily due to the low recognition of MPA (Trost, Rice, & Pfeiffer, 2017). Given the low MVPA values observed in this study and elsewhere (Fairclough et al., 2016; Kim et al., 2017), further calibration work may be necessary to accurately classify MPA from non-processed accelerometer data using the GGIR package. For instance, the auto-calibration procedure within GGIR relies upon the detection of non-movement periods but this method has only been validated in adults across several days (Van Hees et al., 2014), not in children (Hildebrand, Hansen, van Hees, & Ekelund, 2017). Moreover, the auto-calibration method was not used in the laboratory study by Hildebrand et al. (Hildebrand et al., 2014) that proposed the ENMO thresholds. As the auto-calibration method can significantly impact the ENMO values in the lower acceleration range (Hildebrand et al., 2017; Van Hees et al., 2014), future validation studies may wish to develop new ENMO thresholds that use the auto-calibration method.

Attempts have been made to develop conversion equations, known as the Rosetta Stone, to convert estimates of MVPA from widely used cut-points for hip-worn ActiGraph accelerometers into a standardized estimate of MVPA (Brazendale et al., 2016). More recently, a series of Rosetta Stone conversion equations have also been developed to compare estimates of MVPA derived from raw accelerations measured at the wrist and from ActiGraph counts measured at the hip (Brazendale et al., 2018). Applying these Rosetta Stone equations can reduce the differences in estimates of MVPA across studies, but at present they are only available for use with uniaxial hip cut-points. As the authors contend, additional Rosetta Stone conversion equations are needed to compare MVPA estimates from hip-worn ActiGraph VM cut-points which may help researchers compare findings between studies using different processing methods and to pool data.

To the best of our knowledge, this is the first study to document PA estimates from hip-mounted ActiGraph GT3X+ raw and VM count data. As the use of VM cut-points are likely to increase as researchers continue to utilize triaxial accelerometers, we hope our findings draw attention to the limitations of inferring conclusions related to youth activity behaviour between studies using different processing methods. Limitations of this study include the relatively small sample used in this study from one location within Scotland which limits the generalisability of our findings. The lack of criterion measure to estimate time in MPA, VPA and MVPA from VM cut-points can also be considered a limitation. Crucially, estimates of time spent in PA intensities derived from raw accelerations were processed and analysed using open-source procedures which can be used in future studies. This is a strength of this study and allows future studies to compare time spent in MPA, VPA and MVPA with the estimates reported here.

Recent studies have begun to evaluate the classification accuracy of raw and counts based processing methods for classifying SB and PA across standardized activity trials performed in a laboratory (Ellingson et al., 2017; Trost et al., 2017). The results from these studies suggest that recently published cut-points for processed and non-processed accelerometer data in both youth and adults demonstrate poor classification accuracy when compared to criterion measures. Whilst there is some agreement within the literature that the Evenson et al. vertical axis cut-points demonstrate convincing evidence of validity in children (Trost et al., 2011), a similar consensus surrounding VM cut-points or raw processing methods is still lacking. Future work should consider developing a series of Rosetta Stone conversion equations to compare estimates of MVPA derived from raw accelerations and ActiGraph VM counts measured at the hip.

Conclusion

In summary, time spent in activity intensities are not comparable between raw and VM count based data processing methods. The lack of equivalence and poor agreement between raw and VM count processing methods suggest that the two approaches are not comparable. Further calibration work, correction factors and conversion equations may be necessary to facilitate the comparison of findings between studies that process and report raw and VM count activity data.

References

Altman, D. G., & Bland, J. M. (1983). Measurement in Medicine: The Analysis of Method Comparison Studies. *Journal of the Royal Statistical Society. Series D (The Statistician)*, *32*(3), 307–317. https://doi.org/10.2307/2987937

Boddy, L. M., Noonan, R. J., Kim, Y., Rowlands, A. V., Welk, G. J., Knowles, Z. R., & Fairclough, S. J. (2018). Comparability of children's sedentary time estimates derived from wrist worn GENEActiv and hip worn ActiGraph accelerometer thresholds. *Journal of Science and Medicine in Sport*, *0*(0). https://doi.org/10.1016/j.jsams.2018.03.015

Brazendale, K., Beets, M. W., Bornstein, D. B., Moore, J. B., Pate, R. R., Weaver, R. G., … International Children's Accelerometry Database, C. (2016). Equating accelerometer estimates among youth: The Rosetta Stone 2. *J Sci Med Sport*, *19*(3), 242–249. https://doi.org/10.1016/j.jsams.2015.02.006

Brazendale, K., Beets, M. W., Rowlands, A. V., Chandler, J. L., Fairclough, S. J., Boddy, L. M., … Cliff, D. P. (2018). Converting between estimates of moderate-to-vigorous physical activity derived from raw accelerations measured at the wrist and from ActiGraph counts measured at the hip: the Rosetta Stone. *Journal of Sports Sciences*, 1–5. https://doi.org/10.1080/02640414.2018.1470373

Cain, K. L., Sallis, J. F., Conway, T. L., Van Dyck, D., & Calhoon, L. (2013). Using accelerometers in youth physical activity studies: a review of methods. *Journal of Physical Activity & Health*, *10*(3), 437–450.

Choi, L., Liu, Z., Matthews, C. E., & Buchowski, M. S. (2011). Validation of accelerometer wear and nonwear time classification algorithm. *Medicine and*

Science in Sports and Exercise, *43*(2), 357–364. https://doi.org/10.1249/MSS.0b013e3181ed61a3

Cole, T. J., Freeman, J. V., & Preece, M. A. (1995). Body mass index reference curves for the UK, 1990. *Archives of Disease in Childhood*, *73*(1), 25–29.

Cooper, A. R., Goodman, A., Page, A. S., Sherar, L. B., Esliger, D. W., van Sluijs, E. M. F., … Ekelund, U. (2015). Objectively measured physical activity and sedentary time in youth: the International children's accelerometry database (ICAD). *The International Journal of Behavioral Nutrition and Physical Activity*, *12*, 113. https://doi.org/10.1186/s12966-015-0274-5

Crouter, S. E., Horton, M., & Bassett, D. R. (2013). Validity of ActiGraph child-specific equations during various physical activities. *Medicine and Science in Sports and Exercise*, *45*(7), 1403–1409. https://doi.org/10.1249/MSS.0b013e318285f03b

Department of Health. (2011). *Start Active, Stay Active: a report on physical activity for health from the four home countries' Chief Medical Officers*. London.

Dixon, P. M., Saint-Maurice, P. F., Kim, Y., Hibbing, P., Bai, Y., & Welk, G. J. (2018). A Primer on the Use of Equivalence Testing for Evaluating Measurement Agreement. *Medicine and Science in Sports and Exercise*, *50*(4), 837–845. https://doi.org/10.1249/MSS.0000000000001481

Ellingson, L. D., Hibbing, P. R., Kim, Y., Frey-Law, L. A., Saint-Maurice, P. F., & Welk, G. J. (2017). Lab-based validation of different data processing methods for wrist-worn ActiGraph accelerometers in young adults. *Physiological Measurement*, *38*(6), 1045. https://doi.org/10.1088/1361-6579/aa6d00

Eston, R. G., Rowlands, A. V., & Ingledew, D. K. (1998). Validity of heart rate, pedometry, and accelerometry for predicting the energy cost of children's activities. *Journal of Applied Physiology (Bethesda, Md.: 1985)*, *84*(1), 362–371. https://doi.org/10.1152/jappl.1998.84.1.362

Evenson, K. R., Catellier, D. J., Gill, K., Ondrak, K. S., & McMurray, R. G. (2008). Calibration of two objective measures of physical activity for children. *J Sports Sci*, *26*(14), 1557–1565. https://doi.org/10.1080/02640410802334196

Fairclough, S. J., Noonan, R., Rowlands, A. V., Van Hees, V., Knowles, Z., & Boddy, L. M. (2016). Wear Compliance and Activity in Children Wearing Wrist- and Hip-Mounted Accelerometers. *Medicine and Science in Sports and Exercise*, *48*(2), 245–253. https://doi.org/10.1249/MSS.0000000000000771

Grydeland, M., Hansen, B. H., Ried-Larsen, M., Kolle, E., & Anderssen, S. A. (2014). Comparison of three generations of ActiGraph activity monitors under free-living conditions: do they provide comparable assessments of overall physical activity in 9-year old children? *BMC Sports Science, Medicine and Rehabilitation*, *6*, 26. https://doi.org/10.1186/2052-1847-6-26

Hänggi, J. M., Phillips, L. R. S., & Rowlands, A. V. (2013). Validation of the GT3X ActiGraph in children and comparison with the GT1M ActiGraph. *Journal of Science and Medicine in Sport*, *16*(1), 40–44. https://doi.org/10.1016/j.jsams.2012.05.012

Hildebrand, M., Van Hees, V., Hansen, B. H., & Ekelund, U. (2014). Age group comparability of raw accelerometer output from wrist- and hip-worn monitors. *Medicine and Science in Sports and Exercise*, *46*(9), 1816–1824. https://doi.org/10.1249/MSS.0000000000000289

Hildebrand, Maria, Hansen, B. H., van Hees, V. T., & Ekelund, U. (2017). Evaluation of raw acceleration sedentary thresholds in children and adults. *Scandinavian Journal of Medicine & Science in Sports*, *27*(12), 1814–1823. https://doi.org/10.1111/sms.12795

Katzmarzyk, P. T., Barreira, T. V., Broyles, S. T., Champagne, C. M., Chaput, J.-P., Fogelholm, M., … Church, T. S. (2015). Relationship between lifestyle behaviors and obesity in children ages 9–11: Results from a 12-country study. *Obesity*, *23*(8), 1696–1702. https://doi.org/10.1002/oby.21152

Kavouras, S. A., Sarras, S. E., Tsekouras, Y. E., & Sidossis, L. S. (2008). Assessment of energy expenditure in children using the RT3 accelerometer. *Journal of Sports Sciences*, *26*(9), 959–966. https://doi.org/10.1080/02640410801910251

Kim, Y., Crouter, S. E., Lee, J.-M., Dixon, P. M., Gaesser, G. A., & Welk, G. J. (2016). Comparisons of prediction equations for estimating energy expenditure in youth. *Journal of Science and Medicine in Sport*, *19*(1), 35–40. https://doi.org/10.1016/j.jsams.2014.10.002

Kim, Y., Hibbing, P., Saint-Maurice, P. F., Ellingson, L. D., Hennessy, E., Wolff-Hughes, D. L., … Welk, G. J. (2017). Surveillance of Youth Physical Activity and Sedentary Behavior With Wrist Accelerometry. *American Journal of Preventive Medicine*, *52*(6), 872–879. https://doi.org/10.1016/j.amepre.2017.01.012

Logan, G. R. M., Duncan, S., Harris, N. K., Hinckson, E. A., & Schofield, G. (2016). Adolescent physical activity levels: discrepancies with accelerometer data analysis. *Journal of Sports Sciences*, *34*(21), 2047–2053. https://doi.org/10.1080/02640414.2016.1150599

Menai, M., Hees, V. T. van, Elbaz, A., Kivimaki, M., Singh-Manoux, A., & Sabia, S. (2017). Accelerometer assessed moderate-to-vigorous physical activity and successful ageing: results from the Whitehall II study. *Scientific Reports*, *7*, 45772. https://doi.org/10.1038/srep45772

Migueles, J. H., Cadenas-Sanchez, C., Ekelund, U., Nyström, C. D., Mora-Gonzalez, J., Löf, M., … Ortega, F. B. (2017). Accelerometer Data Collection and Processing Criteria to Assess Physical Activity and Other Outcomes: A Systematic Review and Practical Considerations. *Sports Medicine*, *47*(9), 1821–1845. https://doi.org/10.1007/s40279-017-0716-0

Pan, H., & Cole, T. (2010). LMSchartmaker, a program to construct growth references using the LMS method. URL http://www.healthforallchildren.com/shop-base/shop/software/lmsgrowth/ (accessed July 2016). Retrieved from http://healthforallchildren.com/?product=lmsgrowth

Pate, R. R., Stevens, J., Pratt, C., Sallis, J. F., Schmitz, K. H., Webber, L. S., … Young, D. R. (2006). Objectively measured physical activity in sixth-grade girls. *Archives of Pediatrics & Adolescent Medicine*, *160*(12), 1262–1268. https://doi.org/10.1001/archpedi.160.12.1262

Ried-Larsen, M., Brønd, J. C., Brage, S., Hansen, B. H., Grydeland, M., Andersen, L. B., & Møller, N. C. (2012). Mechanical and free living comparisons of four generations of the Actigraph activity monitor. *International Journal of Behavioral Nutrition and Physical Activity*, *9*, 113. https://doi.org/10.1186/1479-5868-9-113

Romanzini, M., Petroski, E. L., Ohara, D., Dourado, A. C., & Reichert, F. F. (2014). Calibration of ActiGraph GT3X, Actical and RT3 accelerometers in adolescents. *European Journal of Sport Science*, *14*(1), 91–99. https://doi.org/10.1080/17461391.2012.732614

Rowlands, A. V., Rennie, K., Kozarski, R., Stanley, R. M., Eston, R. G., Parfitt, G. C., & Olds, T. S. (2014). Children's physical activity assessed with wrist- and hip-worn accelerometers. *Medicine and Science in Sports and Exercise*, *46*(12), 2308–2316. https://doi.org/10.1249/MSS.0000000000000365

Rowlands, Alex V., Cliff, D. P., Fairclough, S. J., Boddy, L. M., Olds, T. S., Parfitt, G., … Beets, M. W. (2016). Moving Forward with Backward Compatibility: Translating Wrist Accelerometer Data. *Medicine and Science in Sports and Exercise*, *48*(11), 2142–2149. https://doi.org/10.1249/MSS.0000000000001015

Rowlands, Alex V., Mirkes, E., Yates, T., Clemes, S., Davies, M., Khunti, K., & L Edwardson, C. (2017). Accelerometer-assessed Physical Activity in Epidemiology: Are Monitors Equivalent? *Medicine & Science in Sports & Exercise*, *50*, 1. https://doi.org/10.1249/MSS.0000000000001435

Rowlands, Alex V., Thomas, P. W. M., Eston, R. G., & Topping, R. (2004). Validation of the RT3 triaxial accelerometer for the assessment of physical activity. *Medicine and Science in Sports and Exercise*, *36*(3), 518–524.

Rowlands, Alex V., Yates, T., Davies, M., Khunti, K., & Edwardson, C. L. (2016). Raw Accelerometer Data Analysis with GGIR R-package: Does Accelerometer Brand Matter? *Medicine and Science in Sports and Exercise*, *48*(10), 1935–1941. https://doi.org/10.1249/MSS.0000000000000978

Sabia, S., Cogranne, P., Hees, V. T. van, Bell, J. A., Elbaz, A., Kivimaki, M., & Singh-Manoux, A. (2015). Physical Activity and Adiposity Markers at Older Ages: Accelerometer Vs Questionnaire Data. *Journal of the American*

Medical Directors Association, *16*(5), 438.e7-438.e13. https://doi.org/10.1016/j.jamda.2015.01.086

Sanders, T., Cliff, D. P., & Lonsdale, C. (2014). Measuring Adolescent Boys' Physical Activity: Bout Length and the Influence of Accelerometer Epoch Length. *PLOS ONE*, *9*(3), e92040. https://doi.org/10.1371/journal.pone.0092040

Santos-Lozano, A., Santín-Medeiros, F., Cardon, G., Torres-Luque, G., Bailón, R., Bergmeir, C., … Garatachea, N. (2013). Actigraph GT3X: validation and determination of physical activity intensity cut points. *International Journal of Sports Medicine*, *35*(11), 975–982. http://dx.doi.org/10.1055/s-0033-1337945

Troiano, R. P., McClain, J. J., Brychta, R. J., & Chen, K. Y. (2014). Evolution of accelerometer methods for physical activity research. *British Journal of Sports Medicine*, *48*(13), 1019–1023. https://doi.org/10.1136/bjsports-2014-093546

Trost, S. G., Loprinzi, P. D., Moore, R. A., & Pfeiffer, K. A. (2011). Comparison of accelerometer cut points for predicting activity intensity in youth. *Medicine and Science in Sports and Exercise*, *43*(7), 1360–1368.

Trost, S. G., Rice, K., & Pfeiffer, K. (2017). Comparison of wrist accelerometer cut-points for classifying physical activity intensity in youth. *Journal of Science and Medicine in Sport*, *20*, e104–e105. https://doi.org/10.1016/j.jsams.2017.01.090

Tudor-Locke, C., Barreira, T. V., Schuna, J. M., Jr., Mire, E. F., Chaput, J. P., Fogelholm, M., … Group, I. R. (2015). Improving wear time compliance with a 24-hour waist-worn accelerometer protocol in the International Study of Childhood Obesity, Lifestyle and the Environment (ISCOLE). *International Journal of Behavioral Nutrition and Physical Activity*, *12*, 11. https://doi.org/10.1186/s12966-015-0172-x

US Department of Health and Human Services. (2008). *Physical Activity Guidlines for Americans. Advisory Committee report 2008. Available from http://www.health.gov/paguidelines/Report/Default.aspx. Accessed 8th August 2011*. Washington DC. Retrieved from www.health.gov/PAGuidlines/committeereport.aspz

Van Hees, V. T., Fang, Z., Langford, J., Assah, F., Mohammad, A., Silva, I. C. M. da, … Brage, S. (2014). Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. *Journal of Applied Physiology*, *117*(7), 738–744. https://doi.org/10.1152/japplphysiol.00421.2014

van Hees, V. T., Gorzelniak, L., Dean Leon, E. C., Eder, M., Pias, M., Taherian, S., … Brage, S. (2013). Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. *PLoS One*, *8*(4), e61691. https://doi.org/10.1371/journal.pone.0061691

World Health Organization. (2010). *Global Recommendations on Physical Activity for Health. Available from http://whqlibdoc.who.int/publications/2010/9789241599979_eng.pdf. Accessed 8th August 2011*. Geneva: World Health Organization.

Table 1. Time spent in activity intensities across three processing methods.

	Romanzini et al.	Hänggi et al.*	Hildebrand et al.
MPA (min.d^{-1})	49.17 (44 to 54)		27.82 (25 to 31)
VPA (min.d^{-1})	42.81 (39 to 47)		6.67 (6 to 7)
MVPA (min.d^{-1})	91.85 (83 to 101)	100.12 (90 to 110)	34.54 (31 to 38)

Data are presented as mean (90% confidence intervals). Abbreviations: MPA = moderate physical activity; VPA = vigorous physical activity; MVPA = moderate to vigorous physical activity. * Cut-points are only available for MVPA.

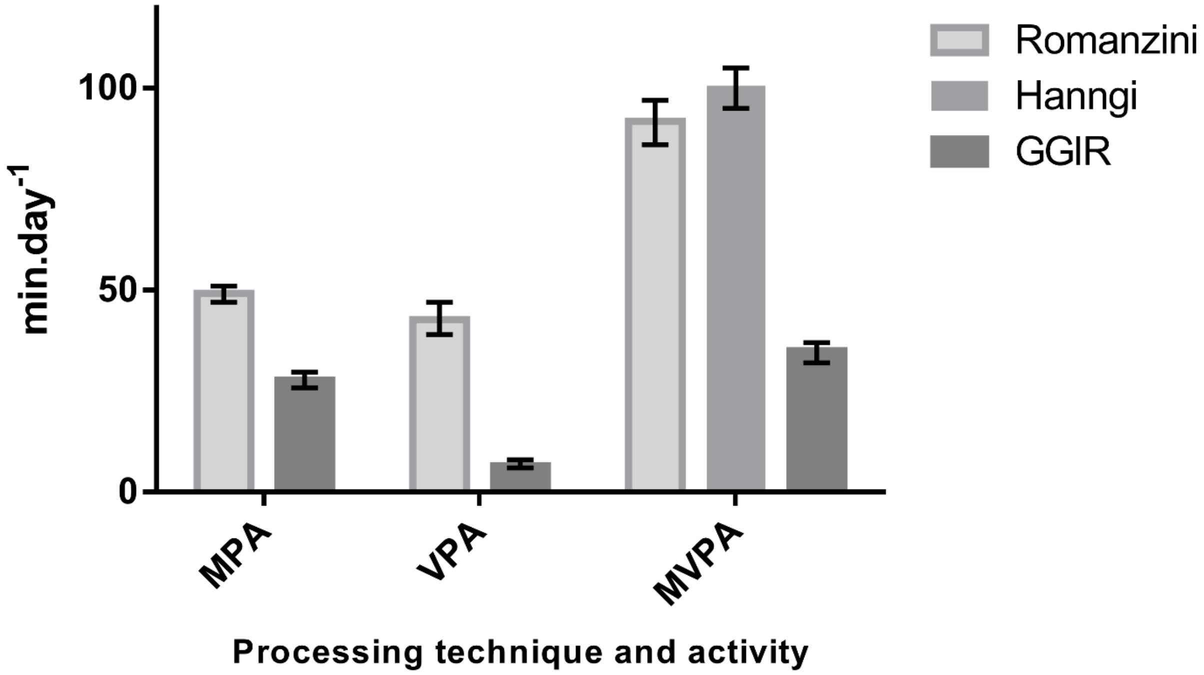

Figure 1. Comparison of three different processing techniques for time recorded in moderate (MPA), vigorous (VPA) and moderate-vigorous (MVPA) physical activity. Data are presented as mean (95%CI).

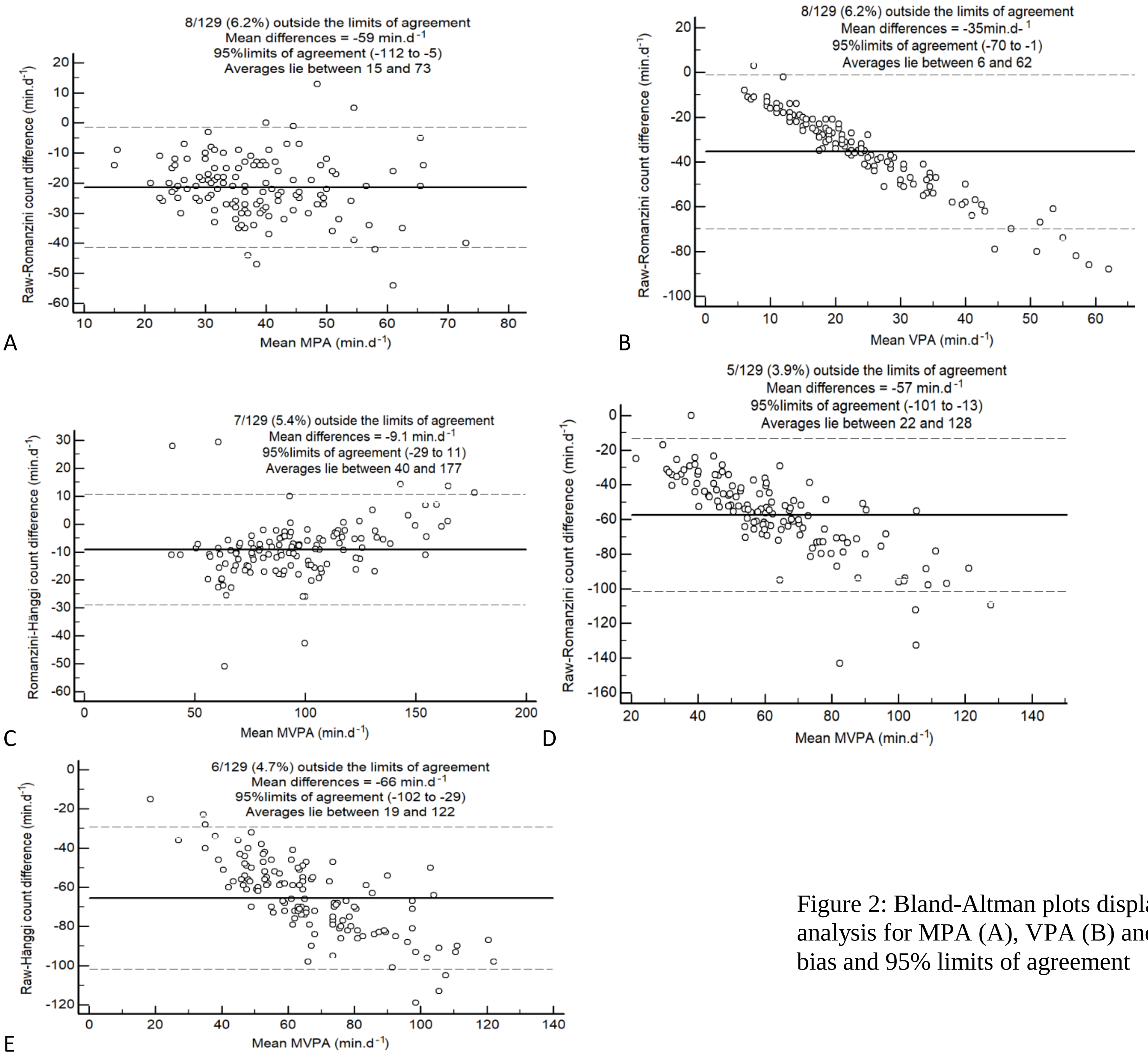

Figure 2: Bland-Altman plots displaying agreement between raw and count data analysis for MPA (A), VPA (B) and MVPA (C-E). Horizontal lines represent mean bias and 95% limits of agreement

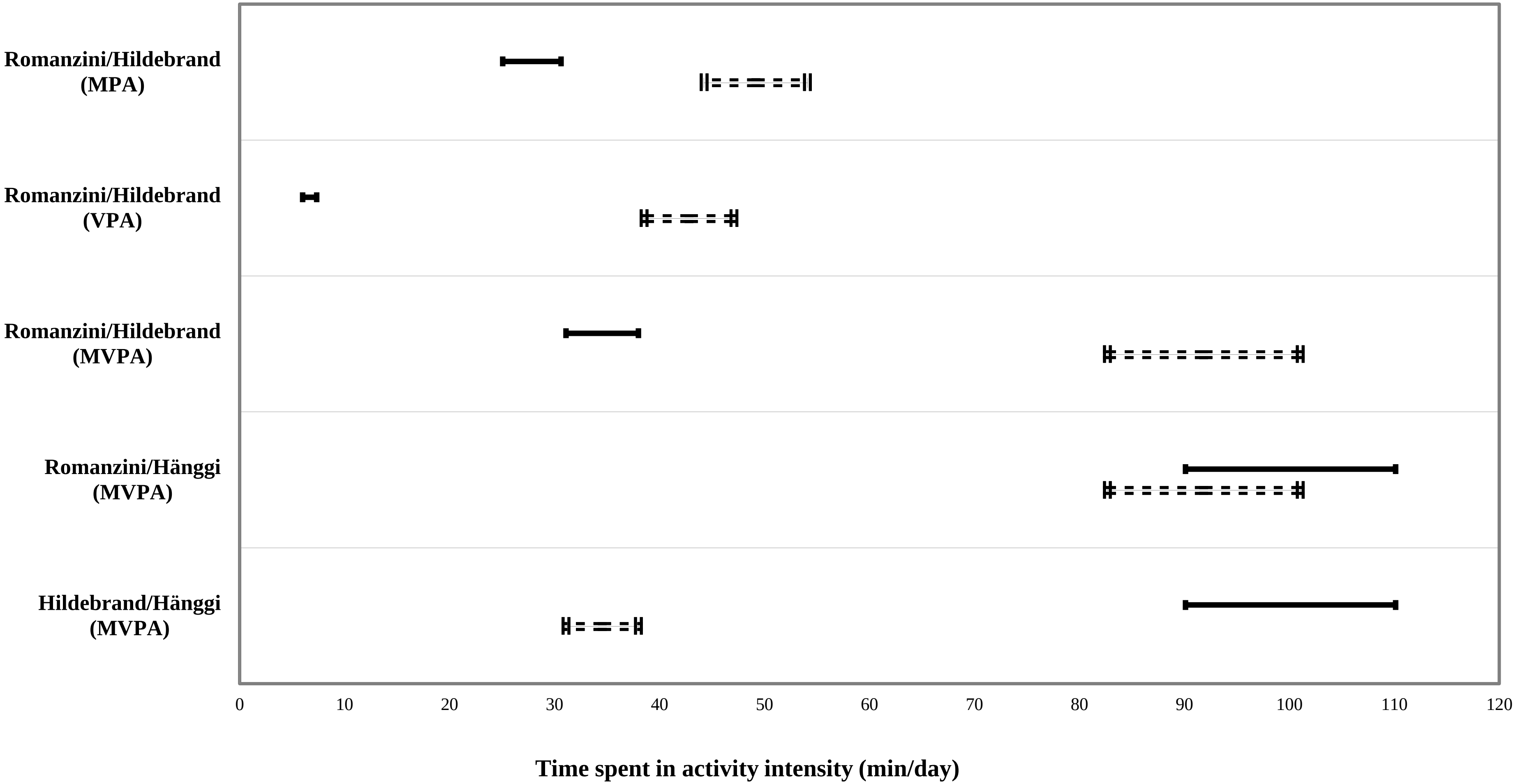

Figure 3. Equivalence between accelerometer data processing methods. Zone of equivalence (dashed lines) and 90% confidence intervals (solid lines) for the reference method*. Abbreviations: MPA = moderate physical activity; VPA = vigorous physical activity; MVPA = moderate-vigorous physical activity. *The first method in each pairing was used as the reference method and are displayed above.